

Published in *Behavioral and Brain Sciences*, 14 (1991), 136-137 as part of a special issue of commentary around a lead article, "The reliability of peer review for manuscripts and grant submissions: A cross-disciplinary investigation." The section runs from pages 119-186.

Does the Need for Agreement Among Reviewers Inhibit the Publication of Controversial Findings?

J. Scott Armstrong
Raymond Hubbard

As Cicchetti indicates, agreement among reviewers is not high. This conclusion is empirically supported by Fiske and Fogg (1990), who reported that two independent reviews of the same papers typically had no critical point in common. Does this imply that journal editors should strive for a high level of reviewer consensus as a criterion for publication? Prior research suggests that such a requirement would inhibit the publication of papers with controversial findings. We summarize this research and report on a survey of editors.

Prior research. Horrobin (1990) suggests that the primary function of peer review should be to identify new and useful findings, that is, to promote the publication of important innovations. This function is typically subordinated to the quality control aspects of peer review, however. The quality control approach looks for agreement among the reviewers. The result, Horrobin claims, is that competent research yielding relatively unimportant findings is more readily accepted for publication.¹ He provides numerous examples of harsh peer review given to important research that presents controversial results.

The popular press often reports difficulties associated with the publication of important research findings. The scanning tunneling microscope (STM) is a case in point. The STM is capable of distinguishing individual atoms and has been hailed as one of the most important inventions of this century. It earned a Nobel Prize in physics for its inventors. Nevertheless, the first attempt to publish the results produced by the STM in 1981 failed because a journal referee found the paper "not interesting enough." (Fisher 1989).

Armstrong (1982c) provides additional examples of lapses in the peer review system, along with summaries of empirical evidence that disconfirming findings about important topics are difficult to publish. Among these, the experimental studies by Goodstein and Brazis (1970) and Mahoney (1977) are of particular interest. They found that reviewers were biased against negative

¹ It is not clear that the quality control function is performed well. About one-third of the papers in biomedical journals were found to contain citation errors, and one-third also incorrectly quoted findings from the literature (Evans et al. (1990). In addition, Hubbard and Armstrong (1990) found that 60% of published replications with extensions in three leading marketing science journals failed to support the original findings.

findings. They rejected these papers on the basis of poor methodology while accepting papers with confirmatory outcomes that used the identical methodology.

Given the above results, one might expect that if editors rely on consensus among reviewers for their publication decisions, few controversial findings will be published. This problem could be especially serious in social science journals. These journals generally have low acceptance rates and their editors may decide to publish only manuscripts with high agreement among reviewers.

A survey of journal editors. To assess how journals treat empirical papers that present controversial findings, we conducted a survey of 20 current or recent editors of American Psychological Association (APA) journals. The two-page questionnaire, together with a stamped, self-addressed return envelope, was mailed out in March 1990. We followed up with phone calls 10 days after the mailing.

Replies were received from 16 of the 20 editors. One question asked: “To the best of your memory, during the last two years of your tenure as editor of an APA journal, did your journal publish one or more papers that were considered to be both controversial and empirical? (That is, papers that presented empirical evidence contradicting the prevailing wisdom.)” Seven editors could recall none.² Four said “yes” and indicated that there was one paper. Three editors replied that there was at least one. Two said that they published several such papers. It seems that controversial empirical papers do get published, but infrequently. Almost half the editors could not recall publishing such papers in the past two years.

We then asked about the peer review for the one published controversial empirical paper that they remembered most clearly. The question was worded: “How did the reviewers respond to this paper?” A five-point scale from “unanimously accepted” to “unanimously rejected” was provided, as well as a “don’t recall” option. One of the nine respondents to this question reported unanimous acceptance, three reported “majority in favor,” four reported “even split,” and one answered “don’t recall.” In response to a question on this published paper’s contribution to the discipline, one editor said “not important,” four said “somewhat important,” and four selected the highest rating, “important.”

The editors were also asked if they had *rejected* any papers that were controversial and empirical. Six of the editors stated that they did not receive such papers, and four said they could not recall any. The six editors who rejected papers with controversial findings did so, they said, because of poor methodology and poor supporting arguments. Of the rejected papers that the editors “remembered most clearly,” only one was “unanimously rejected;” a “majority not in favor” was reported for two, an “even split” for two, and a “majority in favor” for one. Three papers were rated as “not important,” and three as “somewhat important.”

² “Unfortunately,” according to one respondent. Also, in follow-up phone calls, several editors expressed the desire that more such work be submitted.

These results suggest that one can get reviewer agreement on controversial empirical papers. Moreover, most of these papers are published without high levels of reviewer agreement. Apparently, editors do not rely solely on reviewer agreement.

It is interesting that our survey found only two instances of unanimous reviewer agreement for empirical papers with controversial findings. In one case, the recommendation was to reject. In the other, it was to accept. In the case of the accepted manuscript, it should be noted that the editor had invited this submission and had selected reviewers who, he said, were sympathetic to its content.

Our survey indicates that some controversial empirical papers do get published, even when there is disagreement among the reviewers. The willingness of editors to publish such papers is encouraging. On the other hand, 7 of 16 editors could recall no instances of publishing controversial empirical findings. Consequently, we consider some strategies to increase the odds of publishing this type of paper in the next section.

Possible solutions. Some methods that are currently used by journals should help.

1. Some journals' editorial policies allow the author to submit a list of possible referees, one of whom would be selected.
2. Items can be included on structured rating sheets so that reviewers rate the extent to which the findings are controversial. Editors can then give such ratings more weight.
3. Additional reviews can be sought when papers are judged to contain controversial findings. (This strategy was used for only one of the nine published papers and for only one of the six rejected papers in our survey.)
4. Special appeal procedures may help for controversial papers. This might involve other members of the editorial board.
5. Controversial papers can be reviewed initially *without revealing the findings*. This procedure is currently used by the *International Journal of Forecasting*. It has not been used frequently but, when used, it has been beneficial.
6. Provide a section of the journal for "Controversial Findings." The selection of an editor for such a section would indicate the journal's willingness to provide space for such studies. Unfortunately, the one application of this approach that we know (Armstrong 1982b) has produced only one submission, and the findings reported in that submission were not controversial, only the methods were.

Rather than looking for agreement, it might be useful to seek reviewers to act as advocates. This advocacy system would be used for papers that are designated as containing controversial results. A paper could be so designated by the author, the editor, or a reviewer, after which special advocacy procedures would be used. This might include some of the above mentioned suggestions. In addition, one could use more reviewers in an effort to find an advocate. An

advocate could *insist on* publication; a note could be included with the published paper so that reviewers are, in a sense, willing to stake their reputations on the paper.³ Through this note, the readers would receive information about the nature of the acceptance. All referees could be given the opportunity to write peer commentary on the paper. This procedure would greatly increase the likelihood that important papers would be published. The increased effort given to reviewing might also improve quality control.

Conclusions. Controversial empirical papers are expected to receive harsh treatment in peer review, but our survey indicates that such works occasionally get published, sometimes without much peer agreement. More can be done to encourage publication, however. We suggest ways to accomplish this, in particular, the use of an advocacy procedure that explicitly recognizes the need to promote this type of research.

References

- Armstrong, J. S. (1982a), "Barriers to scientific contributions: The author's formula," *Behavioral and Brain Sciences*, 5, 197-199.
- Armstrong, J. S. (1982b), "Is review by peers as fair as it appears?" *Interfaces*, 12, 62-74.
- Armstrong, J. S. (1982c), "Research on scientific journals: Implications for editors and authors," *Journal of Forecasting*, 1, 83-104.
- Fisher, A. (1989), "Seeing atoms," *Popular Science*, 102-07.
- Fiske, D. W. and L. Fogg (1990), "But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments," *American Psychologist*, 45, 591-598.
- Horrobin, D. F. (1990), "The philosophical basis of peer review and the suppression of innovation," *Journal of the American Medical Association*, 263, 1438-1441.
- Mahoney, M. J. (1977), "Publication prejudices: An experimental study of confirmatory bias in the peer review system," *Cognitive Therapy Research*, 1, 161-175.
- McNutt, R. A., A. T. Evans, R. H. Fletcher, and S. W. Fletcher (1990), "The effects of blinding on the quality of peer review," *Journal of the American Medical Association*, 263, 137-176.

³ McNutt *et al.* (1990) found no differences in the quality of reviews based on whether or not they were signed by the reviewer. Also, those who signed the reviews were more likely to recommend acceptance.